Statistics and Computing manuscript No.
(will be inserted by the editor)

Goodness–of–fit tests in semi–linear models

Simos G. Meintanis · Jochen Einbeck

Received: date / Accepted: date

Abstract Specification tests for the error distribution are proposed in semi–linear models, including the partial linear model and additive models. The tests utilize an integrated distance involving the empirical characteristic function of properly estimated residuals. These residuals are obtained from an initial estimation step involving a combination of penalized least squares and smoothing techniques. A bootstrap version of the tests is utilized in order to study the small sample behavior of the procedures in

S. Meintanis

Department of Economics, National and Kapodistrian University of Athens, 8 Pesmazoglou Street, 105 59 Athens, Greece

J. Einbeck (corresponding author)

Department of Mathematical Sciences, Durham University, South Road, Science Laboratories, DH1 3LE, Durham City, UK

Tel.: +44-191-3343125

Fax: +44-191-3343051

E-mail: jochen.einbeck@durham.ac.uk

comparison with more classical approaches. As an example, the tests are applied on some real data sets.

Keywords Semiparametric model · Goodness–of–fit test · Symmetry test · Empirical characteristic function · Bootstrap test

1 Introduction

Suppose that a response variable y is driven by a combination of a linear component and another component which is of unknown functional form. We express the relation between response and predictors through a *semi-linear model*

$$y = \boldsymbol{x}'\boldsymbol{\beta} + g(\boldsymbol{z}) + \sigma\varepsilon \tag{1}$$

where $\boldsymbol{x} = (x_1, \ldots, x_p)'$ and $\boldsymbol{z} = (z_1, \ldots, z_q)'$ denote non-overlapping predictor vectors of dimensions p and q respectively, and where both $\boldsymbol{\beta}$ and $g(\cdot)$ are unknown and have to be estimated from data $\{y_i, \boldsymbol{x}_i, \boldsymbol{z}_i\} \in \mathbb{R}^{1+p+q}$, $i = 1, \ldots, n$. The error ε, which is the actual object of interest of this paper, is assumed to follow an unknown distribution function (DF) F, with $\mathbf{E}(\varepsilon) = 0$ and $\mathbf{E}(\varepsilon^2) = 1$. We assume throughout this paper that the errors $\varepsilon_1, \ldots, \varepsilon_n$ associated to different observations are independent. Important subcases nested in model (1) are the linear model for $q = 0$ and the nonparametric regression model for $p = 0$. Depending on the perspective taken, model (1) has also been referred to as a partial linear model (Speckman, 1988), a semi-parametric model (Hastie & Tibshirani, 1990), or a partial spline model (Wahba, 1984).

We wish to examine two aspects of the corresponding distribution function F of the errors ε: (a) its specific parametric form, i.e. whether F belongs to a specific parametric family of distributions and (b) whether F is symmetric or not. Problem (a) has been

considered in certain popular subcases of model (1) such as the linear regression model and the nonparametric regression model by Jurečková et. al. (2003), Sen et al. (2003), Hušková & Meintanis (2007, 2010), Neumeyer et al. (2006), and Akritas & van Keilegom (2001). For the symmetry problem (b) the reader is referred to Neumeyer & Dette (2007), Hettmansperger et al. (2002), Fan & Gencay (1995) and Dette et al. (2002).

To motivate our procedures we start from Bickel (1982) who provided a general method for constructing asymptotically adaptive and efficient estimates in semiparametric models under certain conditions on the error distribution. However, Schick (1986) points out that Bickel's conditions may not hold for certain error distributions, and yet such adaptive estimates of the non–parametric part in (1) could be feasible, and proceeds to weaken this condition. Subsequent authors study the existence of efficient estimates for $\boldsymbol{\beta}$ with known true error distribution, or under other restrictive assumptions; see for instance Chen (1988), Cuzick (1992) and Schick (1996). As a general message it may be stated that knowledge about the error distribution will ultimately improve statistical analysis for model (1). Also, the linear regression paradigm indicates that if the error distribution is symmetric, efficient adaptive estimation of the regression parameter is always possible; refer to Klaassen & Putter (2005). Additionally, it is well known that certain bootstrap procedures are facilitated considerably under symmetric errors. For extra theoretical and practical information regarding the impact of error–distribution specification on estimation the reader is referred to van der Vaart (1998) and Härdle et al. (2004), respectively.

In this paper we construct testing procedures for the aforementioned null hypotheses (a) and (b) by following the 'Fourier approach' which utilizes the characteristic function (CF). Specifically we consider test statistics which are based on the empirical

CF

$$\varphi_n(t) = \int e^{it\hat{\varepsilon}} dF_n(\hat{\varepsilon}) = \frac{1}{n}\sum_{j=1}^{n} e^{it\hat{\varepsilon}_j},$$

where $F_n(\cdot)$ denotes the empirical DF of the residuals $\hat{\varepsilon}_1, ..., \hat{\varepsilon}_n$, obtained from estimation of the semi–linear model (1). The properties of the estimator F_n of the error DF F have been derived by Müller et al. (2007). In what follows we also explore the small–sample properties of the corresponding classical tests of goodness–of–fit based on this estimator in comparison to the Fourier tests proposed herein. It should be pointed out that generalizing from the classical linear model to the current semiparametric (or to the nonparametric) set–up involves the introduction of an infinite–dimensional 'parameter' $g(\cdot)$ which introduces novel features into the new model. Hence classical methodology, such as analysis of variance and F–tests does not automatically carry over to these more general situations but requires proper modification and new interpretations; see for instance Dette & Neumeyer (2001) and Huang & Davidson (2010). In this connection, and although the proposed test statistics are of similar shape as those considered in a linear and nonparametric regression setup by Hušková & Meintanis (2007, 2010, 2011), the expressions for the limit null distributions given therein do not carry over to our context, since the semiparametric regression setup requires complex estimation routines which inhibit the use of asymptotic theory to a large extent.

The rest of the paper unfolds as follows. In Section 2 we introduce the tests and discuss some aspects of the test statistics. Section 3 deals with the important issue of estimation, while bootstrap versions of the tests are introduced in Section 4 and their behavior is studied by means of Monte Carlo in Section 5. We extend the proposed technique to semi–linear additive models in Section 6. Finally in Section 7 we apply our method to some real data, and summarize our findings in Section 8.

2 Test statistics

Let us write $\varphi_n(t) = C_n(t) + iS_n(t)$, where $C_n(t) = n^{-1}\sum_{j=1}^n \cos(t\hat{\varepsilon}_j)$ is the real part and $S_n(t) = n^{-1}\sum_{j=1}^n \sin(t\hat{\varepsilon}_j)$ is the imaginary part of $\varphi_n(t)$. Likewise, $\phi(t)$ stands for the characteristic function of ε and we denote by $C(t) := \mathbf{E}[\cos(t\varepsilon)]$ and $S(t) := \mathbf{E}[\sin(t\varepsilon)]$, its real and imaginary part, respectively. Also let $\varphi(t) := \varphi_{\boldsymbol{\vartheta}}(t)$ be the CF which corresponds to problem (a) and the null hypothesis $H_0^{(\mathrm{P})} : F \in \mathcal{F}_{\boldsymbol{\vartheta}}$, where $\mathcal{F}_{\boldsymbol{\vartheta}}$ denotes a specific family of distributions, possibly indexed by a parameter $\boldsymbol{\vartheta}$. Then the test statistic for $H_0^{(\mathrm{P})}$ takes the form

$$T_{n,w} = n\int_{-\infty}^{\infty} |\varphi_n(t) - \widehat{\varphi}(t)|^2 w(t)dt, \tag{2}$$

where $\widehat{\varphi}(t) := \varphi_{\widehat{\boldsymbol{\vartheta}}}(t)$ corresponds to estimated parameter $\widehat{\boldsymbol{\vartheta}}$ and $w(t)$ denotes an appropriate weight function the role of which we discuss later. Concerning problem (b) of testing for symmetry, note that $C(t)$ captures the full information on the symmetric component of the error distribution. Hence, the Fourier formulation of the hypothesis of symmetry around the origin is $H_0^{(\mathrm{S})} : S(t) = 0,\ t \in \mathbb{R}$, and the symmetry statistic takes the form

$$S_{n,w} = n\int_{-\infty}^{\infty} S_n^2(t)w(t)dt, \tag{3}$$

where $w(t)$ serves the same purpose as the weight function in (2), but it is not necessarily the same.

The remainder of this section will be devoted to certain expansions corresponding to equations (2) and (3) which will allow us to gain some insight on the test statistics. To this end, we make the following assumptions:

(A1) The weight function satisfies $w(t) = w(-t),\ t \in \mathbb{R}$.

(A2) For some even integer, say $2r$, $\kappa_{2r} := \int_0^{\infty} t^{2r}w(t)dt < \infty$.

(A3) For the same integer as in (A2), $\mu_{2r-1+\delta} < \infty$, for some $0 < \delta \leq 1$, where $\mu_k := \mathbf{E}(|\varepsilon|^k)$.

Based on (A1), it follows that the test statistic in equation (2) admits the representation

$$T_{n,w} = n \int_{-\infty}^{\infty} \left[C_n(t) + S_n(t) - \widehat{C}(t) - \widehat{S}(t) \right]^2 w(t) dt,$$

where $\widehat{C}(t)$ (resp. $\widehat{S}(t)$) denotes the real part (resp. imaginary part) of $\widehat{\varphi}(t)$. Using (A2) and (A3), it follows by Taylor expansions of the trigonometric functions involved in $C_n(\cdot)$ and $S_n(\cdot)$, and by Theorem 2.2.1 of Lukacs (1983) that

$$T_{n,w} = n \sum_{j=1}^{r} \kappa_{2j} f_j(M_1, M_2, ..., M_{2j-1}) + \mathcal{R}_r,$$

where $\mathcal{R}_r := \mathcal{R}_r(\delta, \mu_{2r-1+\delta})$ denotes a remainder. In this equation, $m_k = n^{-1} \sum_{j=1}^{n} \hat{\varepsilon}_j^k$, $k = 1, 2, ...$, are the sample moments and $M_k = m_k - \mathbf{E}(\varepsilon^k | \widehat{\boldsymbol{\vartheta}})$, where $\mathbf{E}(\varepsilon^k | \widehat{\boldsymbol{\vartheta}})$ stands for the moment of order k of $\mathcal{F}_{\boldsymbol{\vartheta}}$ with $\boldsymbol{\vartheta}$ replaced by $\widehat{\boldsymbol{\vartheta}}$. For example if $r = 3$, the 'moment contrasts' f_j, $j = 1, 2, 3$, may be computed by tedious but otherwise straightforward algebra yielding the expansion

$$T_{n,w} = \tag{4}$$
$$n \left[\kappa_2 2 M_1^2 + \kappa_4 \left(\frac{1}{2} M_2^2 - \frac{2}{3} M_1 M_3 \right) + \kappa_6 \left(\frac{1}{30} M_1 M_5 - \frac{1}{12} M_2 M_4 + \frac{1}{18} M_3^2 \right) + \mathcal{R}_3 \right].$$

It is transparent from equation (4) that the CF statistic for testing problem (a) involves moment–matching between the sample moments based on $\hat{\varepsilon}_j$, and the theoretical moments of the hypothesized distribution. In this connection the role of the weight function is to determine the weight κ_{2j} with which each moment equation f_j appears in the test statistic. Under the null hypothesis $H_0^{(\mathrm{P})}$ of course and for large n, each pair of moments (empirical and theoretical) match almost perfectly, and consequently each moment equation f_j and the test statistic itself, should be close to zero. A typical

choice for the weight function is an exponentially decaying weight function, such as $w(t) = e^{-a|t|^b},\ a, b > 0$, which can be easily seen from (4) to yield the limiting values $\lim_{a\to\infty} a^3 T_{n,w} = 4nM_1^2$, and $\lim_{a\to\infty} a^{3/2} T_{n,w} = (\sqrt{\pi}/2)\, nM_1^2$, for $b = 1$ and $b = 2$, respectively.

Using an analogous argument in equation (3) yields the expansion

$$S_{n,w} = 2n\left[\kappa_2 m_1^2 - \frac{2\kappa_4}{1!3!} m_1 m_3 + \frac{\kappa_6}{5!(3!)^2}\left(72 m_1 m_5 + 120 m_3^2\right) + \mathcal{R}_3\right], \qquad (5)$$

(clearly the remainders in equations (4) and (5) are different) which shows that the CF test for symmetry essentially involves odd–order sample moments of the residuals. The limiting values are likewise obtained and correspond exactly to those of the test statistic $T_{n,w}$, but with M_1 being replaced by m_1.

The preceding discussion sheds some light on the criteria based on which the weight function $w(\cdot)$ should be chosen. To begin with, $w(\cdot)$ should be chosen so that the integral figuring in equation (2) can be computed without resorting to numerical integration. Also, among the weight functions ensuring computational simplicity, one should opt for those which secure good power properties. These aspects of $w(\cdot)$ have been discussed by Epps (2005) and Jiménez–Gamero et al. (2009) in the i.i.d. case. Essentially they propose to use a weight function which is proportional to $|\varphi(t)|^2$, where $\varphi(t)$ is the CF under the null hypothesis. This is actually the approach followed in our simulations for testing normality (but there are also other choices that serve the purpose of computational simplicity). Building on this choice, and by introducing an extra parameter a, we use $w(t) = e^{-at^2}$ as a weight function for testing normality. Expansion (4) as well as the limit statistics obtained thereof are illuminating, at least qualitatively, with respect to the value of a. In particular, choosing a large value of a, causes the weight function to decay rapidly, which in turn forces the test statistic to practically 'ignore'

higher order moments, sample and theoretical, and consequently renders its value significantly affected only by few low order moments. In fact, in the limiting case $a \to \infty$ only first order moments have any effect on $T_{n,w}$. On the other hand, choosing a to be too small may cause numerical instability. (Note that for $a = 0$ the test statistic diverges). Hence one may only guess that proper values of a lie somewhere in the interval $0 < a_L < a < a_U < \infty$, between a lower limit a_L and an upper limit a_U, but these values could only be determined empirically via Monte Carlo simulation of the behavior of the test. Otherwise, a more detailed theoretical analysis requires specification of alternative directions away from the null hypothesis; for such an analysis with i.i.d. data and Gram–Charlier alternatives the reader is referred to Epps (1999) and Tenreiro (2009).

3 Estimation in semi–linear models

We recall the setup. We are given data $\{y_i, \boldsymbol{x}_i, \boldsymbol{z}_i\}$, with $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip})'$, and $\boldsymbol{z}_i = (z_{i1}, \ldots, z_{iq})'$, $i = 1, \ldots, n$, and y_i generated according to model (1), i.e.

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + g(\boldsymbol{z}_i) + \sigma\varepsilon_i \tag{6}$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$. The basic problem in calculating the test statistics in (2) and (3) is the estimation of the errors ε_i in (6),

$$\hat{\varepsilon}_i = \frac{y_i - \boldsymbol{x}_i'\hat{\boldsymbol{\beta}} - \hat{g}(\boldsymbol{z}_i)}{\hat{\sigma}},\ i = 1, 2, \ldots, n, \tag{7}$$

which requires the specification of efficient estimators, $\hat{\boldsymbol{\beta}}$ and $\hat{g}(\cdot)$, of the p-dimensional regression parameter $\boldsymbol{\beta}$ and of the nonparametric function $g(\cdot) : \mathbb{R}^q \longrightarrow \mathbb{R}$, respectively, and the use of an appropriate variance estimator $\hat{\sigma}^2$ of σ^2.

A large family of estimators of $\boldsymbol{\beta}$ and $g(\cdot)$ can be derived as solutions to a penalized least squares problem. Denote $\boldsymbol{y} = (y_1, \ldots, y_n)'$, $\boldsymbol{g} = (g(\boldsymbol{z}_1), \ldots, g(\boldsymbol{z}_n))'$, and $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)'$. Given a symmetric $n \times n$ penalty matrix $\boldsymbol{K}$, one aims to minimize

$$Q(\boldsymbol{\beta}, \boldsymbol{g}) = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{g})' (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{g}) + \lambda \boldsymbol{g}' \boldsymbol{K} \boldsymbol{g} \tag{8}$$

w.r.t. $\boldsymbol{\beta}$ and $\boldsymbol{g}$. In the important special case of univariate penalized smoothing splines for $p = 0$ and $q = 1$, the penalty matrix $\boldsymbol{K}$ is constructed such that the penalty term corresponds to $\int g''(t)^2 \, dt$ (see appendix). The solution to (8) is then a natural cubic smoothing spline, i.e. a piecewise cubic polynomial which is connected at the locations of the design points such that the resulting curve is twice continuously differentiable, and has vanishing second and third derivatives at the boundary (Green & Silverman, 1994).

Returning to the general minimization problem (8), we equate $\frac{\partial Q}{\partial \boldsymbol{\beta}}$ and $\frac{\partial Q}{\partial \boldsymbol{g}}$ to zero, yielding

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{X}' (\boldsymbol{y} - \hat{\boldsymbol{g}}) \tag{9}$$

and

$$\hat{\boldsymbol{g}} = (\boldsymbol{I} + \lambda \boldsymbol{K})^{-1} \left(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} \right) \equiv \boldsymbol{S} \left(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} \right) \tag{10}$$

This system of $p + n$ equations is explicitly solvable: by plugging (10) into (9) one has

$$\hat{\boldsymbol{\beta}} = \{\boldsymbol{X}'(\boldsymbol{I} - \boldsymbol{S})\boldsymbol{X}\}^{-1} \boldsymbol{X}'(\boldsymbol{I} - \boldsymbol{S})\boldsymbol{y}. \tag{11}$$

In (10) we have implicitly defined the *smoother matrix* $\boldsymbol{S}$, that is a $n \times n$ matrix $\boldsymbol{S}$ which takes an input vector and produces its smoothed counterpart (see appendix), similar as the hat matrix known from the linear regression model. The analytic solution (11) was already provided in an early paper by Green, Jennison, and Seheult (1985),

who restricted to the case $q = 1$, but mentioned the possibility of extension to bivariate smoothers. For predictors of dimension $q > 1$ nothing is different; the task boils down to constructing an appropriate penalty or smoother matrix and using (10) and (11) (Speckman, 1988). The significance of this result is that no iterative algorithms like backfitting are needed for semi-linear models involving smoothers of arbitrary dimension q. Note also that, given *any symmetric* smoother matrix $\boldsymbol{S}$, (10) tells us that $\boldsymbol{K} \propto (\mathbf{S}^{-} - \boldsymbol{I})$, where $\boldsymbol{S}^{-}$ is a generalized inverse of $\boldsymbol{S}$. Hence, this technique is immediately justified for all symmetric linear smoothers, in the sense that in this case always exists a penalty matrix $\boldsymbol{K}$ such that the resulting estimates can be considered as solutions of a penalized least squares problem. Following Green, Jennison & Seheult (1985) and Speckman (1988), this estimation method can still be used for linear smoothers with asymmetric smoother matrix, though the justification as a penalized least squares solution is lost in this case. The smoother matrices for univariate cubic smoothing splines and local linear smoothers (which are the smoothers used in the simulation study in Section 5) are given in the appendix.

There remains the issue of how to estimate the variance. A natural way of doing this is to compute the residual sum of squares, yielding

$$\hat{\sigma}^2 = \frac{1}{n - df} \sum_{i=1}^{n} \left(y_i - \boldsymbol{x}_i' \hat{\boldsymbol{\beta}} - \hat{g}(\boldsymbol{z}_i) \right)^2 \tag{12}$$

where $df = df_{par} + df_{npar}$ is some measure of the fitted degrees of freedom, consisting of a parametric and nonparametric part. Obviously, $df_{par} = p$, and following the analogue to parametric regression, a straightforward choice is to set $df_{npar} = \mathrm{tr}(\boldsymbol{S})$. A more elaborated solution is obtained by considering the expected residual sum of squares of the smoother, which according to Buja et al. (1989) is given by $\left(n - \mathrm{tr}(2\boldsymbol{S} - \boldsymbol{S}\boldsymbol{S}')\right) \sigma^2 + \mathrm{bias}$. This motivates to use $df_{npar} = \mathrm{tr}(2\boldsymbol{S} - \boldsymbol{S}\boldsymbol{S}')$, which can

be efficiently approximated by

$$df_{npar} \approx 1.25\mathrm{tr}(\boldsymbol{S}) - 0.5 \tag{13}$$

(Hastie & Tibshirani, 1990, Appendix B). We finally note that, in the approach taken above, the degree of freedom corresponding to the intercept parameter is part of df_{npar}, since the intercept is absorbed by the function g, and, hence, by $\boldsymbol{S}$.

4 Bootstrap versions

Due to complicated asymptotics, we develop bootstrap versions of the test statistics in order to actually perform the tests. For the specification null hypothesis $H_0^{(\mathrm{P})}$ we shall restrict the pool of models to simple location–scale families with no extra shape parameters involved. The advantage of considering simple location–scale families is that then the problem is reduced to testing the standard form of this family which is parameter–free. (For this reason the parameter $\boldsymbol{\vartheta}$ could be suppressed.) Specifically, in the context of model (1) the location parameter is set equal to zero, while the scale parameter is estimated and the residuals are standardized accordingly. In principle however, the procedure is applicable to general families of distributions with arbitrary extra parameters, but in this case the issue of these extra parameters should also be addressed during the estimation step; see for instance Hušková and Meintanis (2010). As a result of the preceding discussion, the following procedure is employed in order to compute the critical point of the test for $H_0^{(\mathrm{P})}$:

(i) On the basis of data $\{y_i, \boldsymbol{x}_i, \boldsymbol{z}_i\}$, use (11), (10), and (12) to compute the estimators $(\hat{\boldsymbol{\beta}}, \hat{g}(\cdot), \hat{\sigma})$ and the corresponding residuals $\hat{\varepsilon}_i,\ i = 1, 2, ..., n$.

(ii) Compute the test statistic $T_{n,w} := T_{n,w}(\hat{\varepsilon}_1, ..., \hat{\varepsilon}_n)$.

(iii) Generate i.i.d. observations $\varepsilon_i^*, i = 1, 2, ..., n$, from $\mathcal{F}$ (the hypothesized error distribution under $H_0^{(\mathrm{P})}$), and define the bootstrap observations

$$y_i^* = \boldsymbol{x}_i'\hat{\boldsymbol{\beta}} + \hat{g}(\boldsymbol{z}_i) + \hat{\sigma}\varepsilon_i^*.$$

(iv) Based on $\{y_i^*, \boldsymbol{x}_i, \boldsymbol{z}_i\}$, compute the estimators $(\hat{\boldsymbol{\beta}}^*, \hat{g}^*(\cdot), \hat{\sigma}^*)$ and then based on these estimators compute the corresponding residuals $\hat{\varepsilon}_i^*,\ i = 1, 2, ..., n$, from (7).

(v) Compute the test statistic $T_{n,w}^* := T_{n,w}(\hat{\varepsilon}_1^*, ..., \hat{\varepsilon}_n^*)$.

When steps (iii)–(v) are repeated a number of times, say B, the sampling distribution of $T_{n,w}$ is reproduced, and on the basis of this bootstrap distribution we decide whether the observed value of the test statistic is significant or not.

Likewise, when testing the symmetry null hypothesis $H_0^{(\mathrm{S})}$ with the test statistic $S_{n,w}$, we need only modify step (iii). Specifically step (iii) is modified as follows:

(iii) Define the wild bootstrap residuals

$$\varepsilon_i^* = v_i\hat{\varepsilon}_i,$$

where $v_i,\ i = 1, ..., n$, are i.i.d. observations with $\mathrm{P}(v_i = 1) = \mathrm{P}(v_i = -1) = 1/2$, and define the bootstrap observations

$$y_i^* = \boldsymbol{x}_i'\hat{\boldsymbol{\beta}} + \hat{g}(\boldsymbol{z}_i) + \hat{\sigma}\varepsilon_i^*.$$

For classical statistics, the type of resampling used here has been proposed by Neumeyer et al. (2006) in the case of either linear or nonparametric regression, and was subsequently shown by Hušková & Meintanis (2010) to be asymptotically valid also for CF statistics. On the other hand, the consistency of the wild bootstrap for testing symmetry with classical statistics has been studied by Neumeyer et al. (2005) and Neumeyer & Dette (2007), in the context of linear and nonparametric regression,

respectively, and by Delgado & Escanciano (2007) for dependent data. (An alternative scheme of resampling is the smooth residual bootstrap suggested by Neumeyer, 2009). For analogous work with CF statistics the reader is referred to Hušková & Meintanis (2011). We conclude this section by noting that according to the simulation results in Section 5, the validity of resampling schemes established earlier in the context of linear or nonparametric regression appears to be asymptotically true also in the present context of semi–linear models.

5 Simulations

In this section we investigate the finite–sample behavior of the tests. As an example of the parametric hypothesis $H_0^{(\mathrm{P})}$ we consider testing for normality of the error distribution. Our investigation is carried out by means of a Monte Carlo study. For computational convenience, we use the weight function $w(t) = e^{-at^2}$, and denote the resulting test statistics corresponding to (2) and (3) by $T_{n,a}$ and $S_{n,a}$, respectively. An important aspect of this choice for $w(t)$ is that the integrals figuring in the right–hand sides of equations (2) and (3) can be analytically computed. In particular, $T_{n,a}$ is obtained by replacing $\widehat{\varphi}(t)$ in equation (2) by the normal CF, $e^{-(1/2)t^2}$. Then some straightforward algebra yields

$$T_{n,a} = \frac{1}{n}\sqrt{\frac{\pi}{a}}\left(\sum_{j,k=1}^{n} e^{-(\hat{\varepsilon}_j-\hat{\varepsilon}_k)^2/4a}\right) + n\sqrt{\frac{\pi}{1+a}} - 2\sqrt{\frac{2\pi}{1+2a}}\left(\sum_{j=1}^{n} e^{-\frac{\hat{\varepsilon}_j^2}{(2+4a)}}\right).$$

Likewise, by replacing in equation (3) $w(t)$ by e^{-at^2} yields

$$S_{n,a} = \frac{1}{2n}\sqrt{\frac{\pi}{a}}\sum_{j,k=1}^{n}\left[e^{-(\hat{\varepsilon}_j-\hat{\varepsilon}_k)^2/4a} - e^{-(\hat{\varepsilon}_j+\hat{\varepsilon}_k)^2/4a}\right].$$

We compare our test statistic $T_{n,a}$ with the classical Anderson–Darling (AD) and Cramér–von Mises (CM) statistics, which employ the empirical DF; refer to Section

1. These statistics are routinely employed for testing the parametric null hypothesis $H_0^{(\mathrm{P})}$ with i.i.d. data, as well as for testing for the error distribution in the case of linear or nonparametric regression; see for instance the previously mentioned paper of Neumeyer et al. (2006), as well as the recent works of Heuchenne & van Keilegom (2010) and Neumeyer & van Keilegom (2010). In the context of the semi–linear model (1) however, the behavior of corresponding tests such as the AD and the CM, has not been investigated. Given the order statistics $\hat{\varepsilon}_{(1)} \leq \hat{\varepsilon}_{(2)} \leq ... \leq \hat{\varepsilon}_{(n)}$, these test statistics for normality may be written as (see D'Agostino & Stephens, 1986),

$$T_{\mathrm{AD}} \;=\; -n - \frac{1}{n}\sum_{j=1}^{n}\Big[(2j-1)\log \Phi(\hat{\varepsilon}_{(j)}) + (2(n-j)+1)\log(1-\Phi(\hat{\varepsilon}_{(j)}))\Big]$$

and

$$T_{\mathrm{CM}} \;=\; \frac{1}{12n} + \sum_{j=1}^{n}\left(\Phi(\hat{\varepsilon}_{(j)}) - \frac{2j-1}{2n}\right)^2,$$

respectively, where $\Phi(u)$ denotes the DF of the standard normal distribution. We determine the $p-$values for the AD and CM tests in two different ways: Firstly, based on tabulated values provided in D'Agostino & Stephens (1986, Table 4.9), and secondly, using bootstrap versions of these tests (see also Hušková & Meintanis, 2010). A word of caution is in order: The tabulated $p-$ values correspond to the AD and CM tests as if the errors ε_j are observable, which is clearly not the case. On the contrary expansion (1.2) in Müller et al. (2007) shows that the Kolmogorov–Smirnov type distance between the empirical DF based on ε_j and the empirical DF based on $\hat{\varepsilon}_j$ is not asymptotically negligible to the order of $n^{-1/2}$, and in fact depends on the error density (but not on other aspects of the partial linear model). Consequently, and although our simulation results indicate that at least in the case of normality the difference between the two methods could be considered insignificant, these results certainly do not generalize so as to imply that tabulated $p-$values can be used for other distributions under test.

The performance of the test statistic $S_{n,a}$ is compared to bootstrapped versions of the Kolmogorov–Smirnov and Cramér–von Mises type statistics of Neumeyer & Dette (2007). These statistics are conveniently defined by use of the empirical process

$$D_n(t) = \frac{1}{n}\sum_{j=1}^{n}\left(I(\hat{\varepsilon}_j \le t) - I(-\hat{\varepsilon}_j \le t)\right),$$

as

$$S_{\mathrm{KS}} = \sup_{t\in\mathbb{R}} |D_n(t)| \quad \text{and} \quad S_{\mathrm{CM}} = \int D_n^2(t) dH_n(t),$$

where integration is carried out with respect to the empirical distribution function H_n of $|\hat{\varepsilon}_j|,\ j = 1, ..., n$.

The data are generated from the model

$$y = x + \sin(2\pi z) + \sigma\varepsilon$$

where both x and z are uniformly distributed in the interval $[0, 1]$, and $\sigma = 0.5$. The simulated error distributions are:

(N) Gaussian distribution with mean 0 and standard deviation 1;

(L) Laplace distribution with mean 0 and scale parameter 1;

(SN) Skew-Normal distribution centered at 0, with scale parameter 1 and skew parameter 10;

(SL) Skew-Laplace distribution centered at 0, with scale parameter 1 and skew parameter 3.

The simulated data sets have a sample size of $n = 100$, and we use $B = 200$ bootstrap replicates to carry out each individual test. For each of the error distributions (N), (L), (SN), and (SL), we consider the null hypotheses $H_0^{(\mathrm{P})}$: Normality, and $H_0^{(\mathrm{S})}$: Symmetry. We use the previously developed test statistics $T_{n,a}$ and $S_{n,a}$ for each $a = 1/2$, $a = 1$ and $a = 2$. Specifically, 2000 Monte Carlo replications are generated for

each test problem, and the number of rejections of the corresponding null hypothesis are counted. We repeat the entire procedure using cubic spline smoothers (with constant penalty parameter $\lambda = 5.8 \times 10^{-3}$) and local linear kernel smoothers (with fixed neighborhood size $N = 42$); see appendix A and B for details on the construction of the smoothers. The smoothing parameters were calibrated to produce nonparametric terms corresponding to approximately $\mathrm{tr}(\boldsymbol{S}) = 5$ degrees of freedom. For the estimation of σ, we use (12) and (13).

The percentages of rejections are given in Tables 1 and 2, respectively. One observes that, for underlying Gaussian error, the null hypotheses of symmetry and normality are rejected at a proportion corresponding to the significance level chosen, which is just as it should be. For Laplacian error, the hypothesis of normality is overwhelmingly rejected, while the rejection rate for the symmetry test is slightly above the significance level chosen, but still of an acceptable magnitude. For the skew-normal distribution, both normality and symmetry are clearly rejected for the vast majority of the Monte Carlo replicates. For the skew-Laplace distribution, both hypotheses are rejected at practically all occasions. Throughout all considered testing scenarios, the spline-based smoothers lead to higher test powers (in terms of the proportion of rejection when the null hypothesis is wrong) than the kernel-based smoothers.

Concerning the weight parameter a, it is important to note that for any considered value of a, the Fourier-based tests outperform the Kolmogorov-Smirnov, Anderson–Darling, and Cramér–von Mises statistics both in terms of test power and the accuracy with which the target significance level is met. For the normality test, this holds whether bootstrap or tabulated quantiles were used for the latter. For the test of normality under Laplace error, the power of the test decreased with a, while for the test of symmetry under skew-normal error, the test power increased in tendency with a. Otherwise,

we observed no crucial dependence of the performance of the tests onto the weight parameter a; but in view of the accuracy of the significance level, we would rather recommend to use values of a which are not larger than 1. Finally, we wish to note that, as pointed out by a referee, estimation of σ for the symmetry test is not strictly necessary from a methodological viewpoint, and the bootstrap could be equally carried out using unstandardized residuals. Based on simulation studies which we have carried out, but do not report here for the sake of brevity, we observed indeed higher test powers under this scenario, but at the expense of a greater sensitivity of the method to the choice of a, which in turn impacts negatively on the precision with which the target significance level is met.

6 Semi-linear additive models

Model (1) is attractive from a theoretical point of view, but the q-dimensional surface $g(\boldsymbol{z}) = g(z_1, \ldots, z_q)$ can be difficult to fit in practice due to the so-called curse of dimensionality, which leads to computational problems and to a lack of interpretability in sparse data regions. Often the more realistic option is to combine the individual nonparametric contributions of the components of $\boldsymbol{z}$ additively

$$y = \boldsymbol{x}'\boldsymbol{\beta} + \sum_{j=1}^{q} g_j(z_j) + \sigma\varepsilon \tag{14}$$

or to work with smoothers defined on (usually low-dimensional) non-overlapping subsets $\boldsymbol{t}^{(\ell)}, \ell = 1, \ldots, L$ of $\boldsymbol{z}$ such that $\dot{\bigcup}_\ell \boldsymbol{t}^{(\ell)} = \boldsymbol{z}$ and $\sum_\ell \dim(\boldsymbol{t}^{(\ell)}) = q$, yielding the model

$$y = \boldsymbol{x}'\boldsymbol{\beta} + \sum_{\ell=1}^{L} g_\ell(\boldsymbol{t}^{(\ell)}) + \sigma\varepsilon. \tag{15}$$

Table 1 Percentage of rejection of the null hypothesis $H_0^{(\mathrm{P})}$ (Normality) for four different true error distributions, and three different choices of a. Top: using smoothing splines; bottom: using local linear kernel smoothers. The suffix t indicates the tabulated versions of the AD and CM tests, while all other columns refer to bootstrapped tests.

Splines		$a = 1/2$	$a = 1$	$a = 2$	AD_t	AD	CM_t	CM
(N)	$\alpha = 0.05$	5.3	5.1	5.0	4.7	4.8	5.5	5.3
	$\alpha = 0.10$	10.6	10.8	10.5	10.8	10.5	11.0	11.0
(L)	$\alpha = 0.05$	79.0	77.6	72.9	73.3	72.4	71.2	70.6
	$\alpha = 0.10$	86.9	86.0	82.5	80.7	80.5	78.9	78.7
(SN)	$\alpha = 0.05$	84.6	87.0	85.4	83.1	82.0	77.6	76.9
	$\alpha = 0.10$	91.6	92.2	91.9	89.5	89.3	85.9	85.6
(SL)	$\alpha = 0.05$	100.0	100.0	100.0	100.0	100.0	99.9	99.9
	$\alpha = 0.10$	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Kernels		$a = 1/2$	$a = 1$	$a = 2$	AD_t	AD	CM_t	CM
(N)	$\alpha = 0.05$	4.5	4.6	4.4	5.0	5.1	5.3	5.1
	$\alpha = 0.10$	10.4	9.7	9.3	10.8	10.1	10.6	10.9
(L)	$\alpha = 0.05$	77.4	75.1	69.2	70.3	70.4	68.5	67.9
	$\alpha = 0.10$	86.2	84.8	80.0	79.4	79.1	77.2	77.3
(SN)	$\alpha = 0.05$	80.4	81.8	78.5	78.3	77.6	73.6	73.1
	$\alpha = 0.10$	88.0	89.3	86.3	87.3	86.8	82.1	82.0
(SL)	$\alpha = 0.05$	99.9	100.0	100.0	99.9	99.9	99.9	99.9
	$\alpha = 0.10$	100.0	100.0	100.0	100.0	100.0	100.0	100.0

We refer to models of type (14) and (15) as *semi-linear additive models*. The intercept term, say β_0, needs now to be incorporated into the parametric part $\boldsymbol{x}'\boldsymbol{\beta}$ as identifiability problems arise otherwise (Fahrmeir & Tutz, 2001). Obviously, (15) covers (14), and also covers the so-called *additive model* where $\boldsymbol{x}'\boldsymbol{\beta} = \beta_0$ (Hastie & Tibshirani, 1990).

Table 2 Percentage of rejection of the null hypothesis $H_0^{(\mathrm{S})}$ (symmetry) for four different true error distributions, and three different choices of a. Top: using smoothing splines; bottom: using local linear kernel smoothers.

Splines		$a = 1/2$	$a = 1$	$a = 2$	KS	CM
(N)	$\alpha = 0.05$	5.2	5.2	4.5	4.2	5.7
	$\alpha = 0.10$	10.0	10.0	9.0	8.5	11.4
(L)	$\alpha = 0.05$	6.2	6.3	6.8	7.3	7.7
	$\alpha = 0.10$	11.7	12.0	12.2	12.4	13.0
(SN)	$\alpha = 0.05$	83.1	86.7	87.9	65.6	74.2
	$\alpha = 0.10$	89.6	92.2	93.2	78.2	82.9
(SL)	$\alpha = 0.05$	100.0	100.0	100.0	99.3	99.9
	$\alpha = 0.10$	100.0	100.0	100.0	99.8	100.0
Kernels		$a = 1/2$	$a = 1$	$a = 2$	KS	CM
(N)	$\alpha = 0.05$	5.6	5.6	4.7	4.2	6.3
	$\alpha = 0.10$	10.1	9.8	9.3	9.1	11.4
(L)	$\alpha = 0.05$	6.3	6.4	6.1	6.3	7.7
	$\alpha = 0.10$	11.9	12.2	11.7	11.3	13.5
(SN)	$\alpha = 0.05$	78.2	80.7	79.5	59.7	67.8
	$\alpha = 0.10$	85.7	87.9	87.2	71.4	77.7
(SL)	$\alpha = 0.05$	100.0	100.0	99.9	98.6	99.8
	$\alpha = 0.10$	100.0	100.0	100.0	99.6	100.0

The testing procedure that we have proposed in Sections 2 and 4 extends straightforwardly to this setting. However, the estimation of parameters and smooth terms is slightly more involved, for which reason we give the corresponding formulas explicitly below.

In terms of (15), the minimization problem takes the shape

$$Q(\boldsymbol{\beta}, \boldsymbol{g}_1, \ldots, \boldsymbol{g}_L) = \big(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \textstyle\sum_\ell \boldsymbol{g}_\ell\big)' \big(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \sum_\ell \boldsymbol{g}_\ell\big) + \sum_\ell \lambda_\ell \boldsymbol{g}_\ell' \boldsymbol{K}_\ell \boldsymbol{g}_\ell$$

where $\boldsymbol{g}_\ell = (g_\ell(\boldsymbol{t}_1^{(\ell)}), \ldots, g_\ell(\boldsymbol{t}_n^{(\ell)}))'$, with $\boldsymbol{t}_i^\ell$ being the corresponding ℓ-th subset of $\boldsymbol{z}_i$. The matrices $\boldsymbol{K}_\ell$ are $n \times n$ penalty matrices with associated smoother matrices $\boldsymbol{S}_\ell$, $\ell = 1, \ldots, L$. Equating $\frac{\partial Q}{\partial \boldsymbol{\beta}}$ and $\frac{\partial Q}{\partial \boldsymbol{g}_\ell}$ to zero, one finds that the equivalent to (9) is given by

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\left(\boldsymbol{y} - \textstyle\sum_\ell \hat{\boldsymbol{g}}_\ell\right), \tag{16}$$

while that of (10) is

$$\hat{\boldsymbol{g}}_\ell = \boldsymbol{S}_\ell(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - \textstyle\sum_{k\neq\ell} \hat{\boldsymbol{g}}_k). \tag{17}$$

However, it turns out that the resulting system of $p + nL$ equations is *not* explicitly solvable any more. Hence, one has to resort to the backfitting algorithm, which was introduced and studied in detail in the context of the additive model by Buja et al. (1989). Adapted to the semi–linear additive model for general q, the backfitting algorithm reads as

(i) Initialize: $\boldsymbol{\beta} = \boldsymbol{\beta}_0$, $\boldsymbol{g}_\ell = \boldsymbol{g}_\ell^0$, $\ell = 1 \ldots, L$.

(ii) Estimate $\hat{\boldsymbol{\beta}}$ according to (16).

(iii) For $\ell = 1, \ldots, L$, update $\hat{\boldsymbol{g}}_\ell$ according to (17).

(iv) Cycle (ii) and (iii) until the individual functions and parameters do not change.

A variant of this is to separate the nonparametric part into a parametric ('projection') and the remaining nonparametric ('shrinking') part, and estimate the projection part together with the parametric part in step (ii). This method has several computational advantages. The results are exactly the same as for the original backfitting algorithm only for a subclass of symmetric linear smoothers which includes smoothing splines (see Hastie & Tibshirani (1990), p. 124 ff., for details). This variant, which is implemented in R function `gam` (Hastie, 1992), is used in the oceanographic data example in Section 7.

7 Real data examples

We firstly consider a data set which was used in Ruppert et al. (2003) to illustrate the performance of semi-parametric regression estimates. We are given 84 observations from an experiment involving the production of white Spanish onions at two South Australian locations. The covariates available are the areal `density` of plants measured in plants per square meter, and an indicator variable `location` taking the value 1 if the measurement was taken at 'Purnong Landing'. Estimation of a linear model

$$\log(\texttt{yield}) = \beta_0 + \beta_1 \texttt{location} + \beta_2 \texttt{density}$$

yields the fitted straight lines in Figure 1. Using the bootstrap technique introduced in Section 4, one obtains a $p-$value of 0.02 for testing normality and a $p-$value of 0.16 for testing symmetry, which gives some evidence that the fitted linear model is not adequate. Fitting now a semi-linear model

$$\log(\texttt{yield}) = \beta_1 \texttt{location} + g(\texttt{density}),$$

these $p-$ values change to 0.79 and 0.84, respectively, indicating an improved goodness-of-fit when accounting for the nonlinear dependence of `log(yield)` on `density`. We have used here a local linear smoother with neighborhood parameter $N = 39$, again corresponding to roughly $\mathrm{tr}(\boldsymbol{S}) = 5$ degrees of freedom. Of course splines could be used here equally well, but, as there are tied `density` values, this would require appropriate grouping and weighting before the estimation techniques outlined in Section 3 could be applied (Hastie & Tibshirani, 1990, p. 74).

Secondly, we consider oceanographic data retrieved from the World Ocean Database by Powell (2009). The data were collected in the North Atlantic by the German vessel "Gauss", yielding $n = 643$ measurements on several variables, including the water

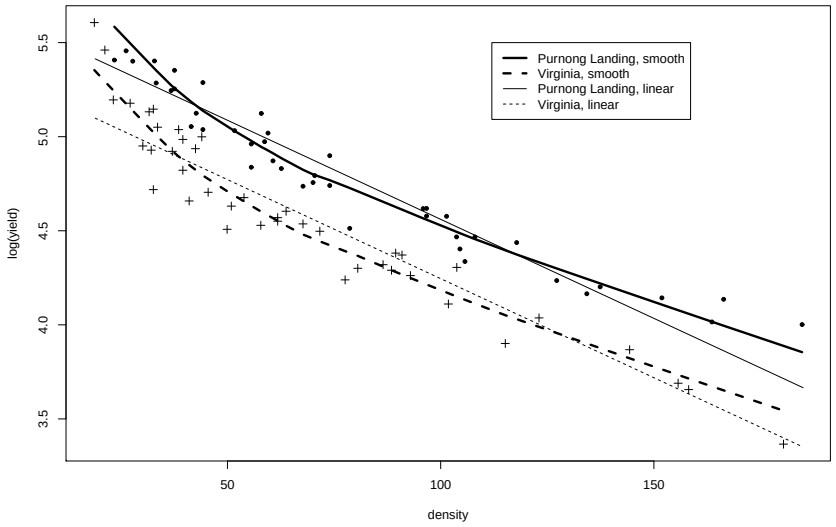

Fig. 1 Logarithm of onion yield vs. areal density of plants, measured at two locations in Southern Australia. The straight (thin) lines correspond to the linear model fit, while the smooth curves are obtained by fitting a semi-parametric model.

temperature in degrees Celsius (this serves as the response), the salinity of the water (measured in the Practical Salinity Scale, PSS), the oxygen content in millimeters per litre of water, and the depth under the surface (in meters) at which the measurement was taken. Fitting an additive model (A) of type (14) with an intercept and $q = 3$ non-parametric terms (via cubic smoothing splines with $\mathrm{tr}(\boldsymbol{S}_\ell) - 1 = 6$ df per model term, $\ell = 1, 2, 3$), yields the three fitted functions depicted in Figure 2. The corresponding goodness-of-fit tests deliver a $p-$value of 0.00 for the symmetry test and 0.00 for the normality test,[1] so both null hypotheses are clearly rejected.

The first of the three smooth curves seems to suggest that the impact of salinity onto temperature could be rather linear than nonlinear. Hence, it seems a natural idea

[1] To be precise, a p–value of 0.00 obtained in this manner, using $B = 200$, means $p < 0.005$.

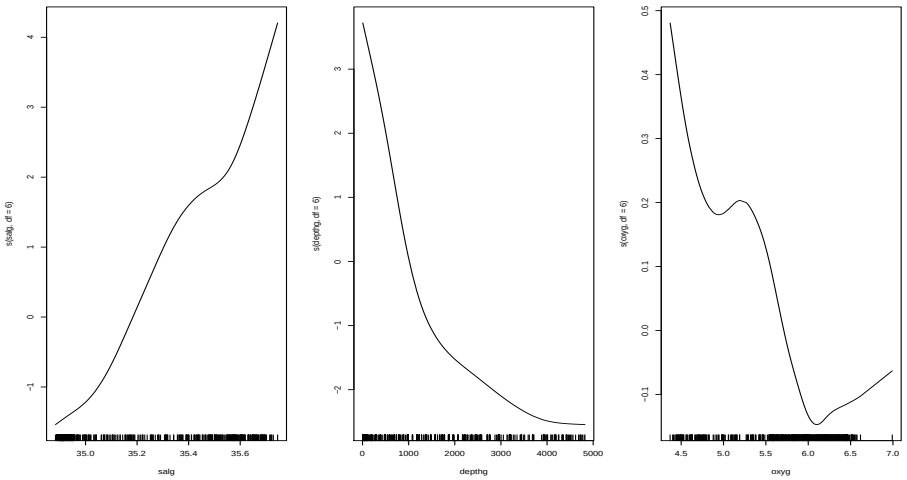

Fig. 2 Fitted nonparametric terms for salinity, water depth, and oxygen.

to replace this nonparametric term by a linear term, and observe whether this has implications for the goodness-of-fit. This gives a semi-linear additive model (B) of type (14) with $p = 2$ (including the intercept) and $q = 2$. Interestingly, after having replaced the nonlinear by a linear term, the $p-$value for symmetry increases to 0.07 (with that one for normality remaining at 0.00). It seems plausible that the way that oxygen content influences temperature depends on the water depth. We therefore consider a model (C) featuring a bivariate "surface smoother" (Hastie, 1992) for oxygen and water depth, and a linear term for salinity, which is a semiparametric model of type (1) with $p = 1$ (now excluding the intercept) and $q = 2$, where $\boldsymbol{z} = (\text{water depth}, \text{oxygen})$. The goodness-of-fit tests for this semi-linear model give a $p-$value of 0.13 for the symmetry and 0.00 for the test of normality, indicating a symmetrical, though non-normal, behavior of the residuals. These results are qualitatively confirmed by looking at histograms and Gaussian probability plots (QQ-Plots) of the corresponding residual

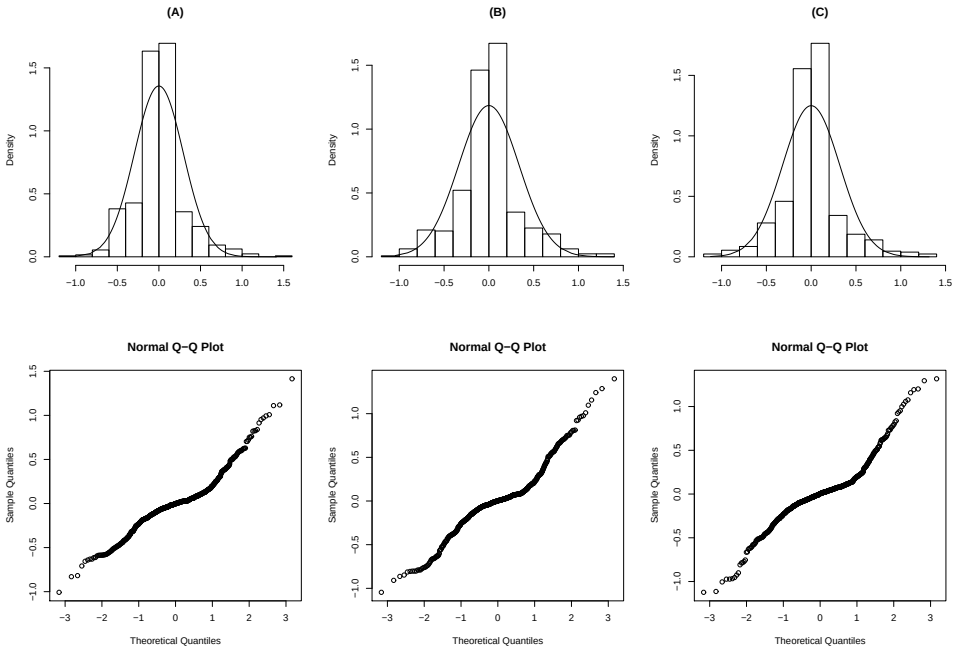

Fig. 3 Histograms (with overlaid normal densities) and QQ plots for the residuals of the additive model (A), the semi-linear additive model (B), and the semi-linear model (C) for the oceanographic data. Note the improving symmetry from left to right, corresponding to the $p-$values 0.00, 0.07, and 0.13, respectively, of the symmetry test.

distributions, which are provided in Figure 3. For ease of interpretation, parametric estimates of Gaussian densities are overlaid over the histograms. One observes that all distributions show deviations from normality, in particular around the peaks, with that one based on the interaction model (C) being more symmetrically distributed than the others.

8 Discussion

The purpose of this work is (i) to develop Fourier–type goodness–of–fit procedures for semi–linear models based on the empirical characteristic function, and (ii) to compare these procedures with classical procedures based on the empirical distribution function. In doing so we have considered different estimators of the components of the semi–linear model, and have studied original and bootstrap versions of the tests. The general messages from our simulation results are that (i) all methods recover the nominal size of the tests to a satisfactory degree, (ii) splines rather than kernels lead to somewhat higher power, (iii) bootstrap and original versions result in almost indistinguishable rejection rates, and that (iv) Fourier–type tests are more powerful than classical tests, though not by a wide margin.

As noted above, the goodness–of–fit tests proposed have been implemented by using spline– and kernel– based estimators for the nonparametric part. The vehicle for estimation that we have used builds on normal equations motivated originally in the context of penalized least squares regression (Green, Jennison, & Seheult, 1985), and developed further in particular by Hastie & Tibshirani (1990). Though we have investigated the performance of our testing routines only for this particular way of estimation, there is no apparent reason as to why these tests couldn't be applied onto models fitted through other semi-parametric regression techniques, such as the direct kernel approach by Robinson (1988) or the mixed model approach by Ruppert et al. (2003). Preliminary investigations using these techniques led to encouraging results, so we tentatively recommend the proposed tests beyond the framework of the estimation methods considered here. Furthermore, it is also clear that the test procedures for the null hypothesis $H_0^{(\mathrm{P})}$ proposed herein do not serve the sole purpose of test-

ing normality, but can readily be applied to any other error distribution under test. This is particularly important given the fact that applied workers, particularly in the area of empirical finance, have long rejected the assumption of normality and operate under distributions that are both asymmetric and heavy tailed. In this connection, and although as mentioned above there seems to be no apparent gain in power, bootstrap quantiles are to be preferred over tabulated ones as they are readily operational regardless of the method of estimation and the postulated error distribution.

We close with a word of caution: The notion of *goodness-of-fit* advocated here refers to certain aspects of the error distribution, and therefore it should not be confused with that of *significance* of parameters or smooth terms. Hence, rather than considering it as a competitor to F–tests, our method may serve as a vehicle to justify or discard the application of the latter: If the null hypothesis of normality (of the smaller model) is rejected, then the application of the F-test is not justified, as it uses the assumption of Gaussian errors under the null hypothesis that the smaller model is correct. In fact, when carrying out the appropriate F–test comparing the linear (B) with the nonparametric (A) term for salinity, it turns out that model (B) is clearly rejected in favor of (A)[2]; but as shown in Section 7, the application of the F-test itself is not endorsed by the normality test.

Appendix: Linear smoothers and smoother matrices

Suppose we are given data $(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_n, y_n)$ generated from a model of type

$$y_i = g(\boldsymbol{x}_i) + \sigma\varepsilon_i$$

[2] In fact, one can even argue that the kink in the smooth term for salinity is biologically plausible, see Powell (2009) for details.

where ε_i is noise with mean zero and unit variance. Given a nonparametric smoother, i.e. a twice continuously differentiable function $\hat{g} : \mathbb{R}^p \longrightarrow \mathbb{R}$, the *smoother matrix* $\boldsymbol{S}$ is defined as the $n \times n$ matrix which maps a vector of observed responses $\boldsymbol{y} = (y_1, \ldots, y_n)'$ to their fitted (smoothed) values $\hat{\boldsymbol{y}} = (\hat{y}_1, \ldots, \hat{y}_n)'$, where $\hat{y}_i = \hat{g}(\boldsymbol{x}_i)$. If such a smoother matrix exists which does not depend on $\boldsymbol{y}$, then the smoother is called a *linear smoother* (Buja et al., 1989), and one has

$$\hat{\boldsymbol{y}} = \boldsymbol{S}\boldsymbol{y}.$$

If $\boldsymbol{S}$ is symmetric, $\hat{g}$ is called a *symmetric linear smoother*. Let $\boldsymbol{g} = (g(\boldsymbol{x}_1), \ldots, g(\boldsymbol{x}_n))'$. Symmetric linear smoothers can be considered as minimizers of the penalized least squares problem

$$(\boldsymbol{y} - \boldsymbol{g})'(\boldsymbol{y} - \boldsymbol{g}) + \lambda \boldsymbol{g}' \boldsymbol{K} \boldsymbol{g}$$

where $\boldsymbol{K} \propto (\mathbf{S}^- - \boldsymbol{I})$, with $\boldsymbol{S}^-$ being a generalized inverse of $\boldsymbol{S}$ (Hastie & Tibshirani, 1990, p. 110).

Smoother matrices for two important special cases involving univariate predictors $x_1, \ldots, x_n \in \mathbb{R}$ are provided below. A quite comprehensive overview of other smoothers and their associated smoother matrices is provided in Buja et al. (1989).

A. Cubic smoothing splines. Assume we have an ordering of predictors such that $x_1 < x_2 < \ldots < x_n$, and let $d_i = x_{i+1} - x_i$. The penalty matrix is given by

$\boldsymbol{K} = \boldsymbol{D}'\boldsymbol{C}^{-1}\boldsymbol{D}$, where

$$\boldsymbol{D} = \begin{pmatrix} \frac{1}{d_1} & -\left(\frac{1}{d_1}+\frac{1}{d_2}\right) & \frac{1}{d_2} & & & \\ & \frac{1}{d_2} & -\left(\frac{1}{d_2}+\frac{1}{d_3}\right) & \frac{1}{d_3} & & \\ & & \ddots & \ddots & \ddots & \\ & & & \frac{1}{d_{n-2}} & -\left(\frac{1}{d_{n-2}}+\frac{1}{d_{n-1}}\right) & \frac{1}{d_{n-1}} \end{pmatrix}$$

is an $n-2 \times n$ upper-tridiagonal matrix and

$$\boldsymbol{C} = \frac{1}{6}\begin{pmatrix} 2(d_1+d_2) & d_2 & & & \\ d_2 & 2(d_2+d_3) & d_3 & & \\ & \ddots & \ddots & \ddots & \\ & & & d_{n-2} & \\ & & & d_{n-2} & 2(d_{n-2}+d_{n-1}) \end{pmatrix}$$

is an $n-2 \times n-2$ tridiagonal symmetric matrix (Green & Silverman, 1994, p. 12f; Fahrmeir & Tutz, 2001, p. 181f). For fixed smoothing parameter λ, the smoother matrix is then obtained by taking $\boldsymbol{S} = (\boldsymbol{I} + \lambda\boldsymbol{K})^{-1}$.

B. Local linear smoothers. Denote $K : \mathbb{R} \longrightarrow \mathbb{R}^+$ a symmetric kernel function. The smoother matrix $\boldsymbol{S} = (s_{ij})_{1\le i\le n, 1\le j\le n}$ is specified by

$$s_{ij} = \left(S_{n,0}(x_j)S_{n,2}(x_j) - S_{n,1}^2(x_j)\right)^{-1} K\left(\frac{x_i - x_j}{h(x_j)}\right)\left(S_{n,2}(x_j) - (x_i - x_j)S_{n,1}(x_j)\right)$$

with bandwidths $h(x_j) \in \mathbb{R}^+$ and

$$S_{n,\ell}(x) = \sum_{i=1}^{n} K\left(\frac{x_i - x}{h(x)}\right)(x_i - x)^{\ell}.$$

Two important subcases are the use of a *global bandwidth* $h(x) \equiv h$, and the use of N *nearest neighbors*, in which case $h(x)$ is the distance to the $N-$th nearest neighbor to

x. In the former case it is common to work with a Gaussian or Epanechnikov kernel K (Fan & Gijbels, 1996), while in the latter case commonly a tricube weight function $K(t) = \frac{70}{81}(1 - |t|^3)^3 I_{[-1,1]}(t)$ is used (Cleveland, 1979). In either case, this smoother matrix is asymmetric, implying that there is no exact representation in form of a penalty matrix $\boldsymbol{K}$. The simulations performed for Table 1 and 2 use the variant based on nearest neighbors and the tricube kernel.

Acknowledgements The authors wish to thank B. Powell for providing the oceanographic data set and for sharing his insight into biological and statistical aspects of it. Part of this study was conducted during the first author's visit to the Department of Mathematical Sciences, University of Durham. SGM wishes to sincerely thank the Department for its hospitality and support.

References

1. Akritas, M., and van Keilegom, I.: Non-parametric estimation of the residual distribution. *Scand. J. Statist.*, 28, 549–567 (2001)
2. Bickel, P.J.: On adaptive estimation. *Ann. Statist.*, 10, 647–671 (1982)
3. Buja, A., Hastie, T., and Tibshirani, R.: Linear smoothers and additive models (with discussion). *Ann. Statist.*, 17, 453–555 (1989)
4. Chen, H.: Convergence rates for parametric components in a partly linear model. *Ann. Statist.*, 16, 136–146 (1988)
5. Cleveland, W. S.: Robust locally weighted regression and smoothing scatterplots. *J. Amer. Statist. Assoc.*, 74, 829–836 (1979)
6. Cuzick, J.: Semiparametric additive regression. *J. Roy. Statist. Soc. B*, 54, 831–843 (1992)
7. D'Agostino, R. and Stephens, M.: *Goodness-of-fit techniques.* Marcel Dekker, Inc., New York (1986)
8. Delgado, M.A., and Escanciano, J.C.: Nonparametric tests for conditional symmetry in dynamic models. *J. Econometr.*, 141, 652–682 (2007)

9. Dette, H., Kusi–Appiah, S., and Neumeyer, N.: Testing symmetry in nonparametric regression models. *J. Nonparam. Statist.*, 14, 477–494 (2002)

10. Dette, H., and Neumeyer, N.: Nonparametric analysis of covariance. *Ann. Statist.*, 29, 1361–1400 (2001)

11. Epps, T.W.: Limiting behavior of the ICF test for normality under Gram–Charlier alternatives. *Statist. Probab. Lett.*, 42, 175–184 (1999)

12. Epps, T.W.: Tests for location–scale families based on the empirical characteristic function. *Metrika*, 62, 99–114 (2005)

13. Fahrmeir, L. and Tutz, G.: *Multivariate Statistical Modelling based on Generalized Linear Models.* Springer Series in Statistics, Springer (2001)

14. Fan, Y., and Gencay, R.: A consistent nonparametric test of symmetry in linear regression models. *J. Amer. Statist. Assoc.*, 90, 551–557 (1995)

15. Fan, Y. and Gijbels, I.: *Local polynomial modelling and its applications.* Monographs on Statistics and Probability. Chapman & Hall (1996)

16. Green, P., Jennison, C., and Seheult, A.: Analysis of field experiments by least squares smoothing. *J. Roy. Statist. Soc. B*, 47, 299–315 (1985)

17. Green, P.J. and Silverman, B.: *Nonparametric Regression and Generalized Linear Models.* Monographs on Statistics and Probability. Chapman & Hall (1994)

18. Härdle, W., Müller, M., Sperlich, S. and Werwatz, A.: *Nonparametric and Semiparametric Models.* Springer Series in Statistics. Springer (2004)

19. Hastie, T.J.: Generalized Additive Models. In: Chambers, J., and Hastie, T. (Eds). *Statistical Models in S.* Chapman & Hall (1992)

20. Hastie, T.J. and Tibshirani, R.J.: *Generalized Additive Models.* Monographs on Statistics and Applied Probability. Chapman & Hall (1990)

21. Hettmansperger, T.P., McKean, J.W., and Sheather, S.J.: Finite sample performance of tests for symmetry of the errors in a linear model. *J. Statist. Comput. Simul.*, 72, 863–879 (2002)

22. Heuchenne, C., and van Keilegom, I.: Goodness–of–fit tests for the error distribution in nonparametric regression. *Comput. Statist. Dat. Anal.*, 54, 1942–1951 (2010)

23. Huang, L.S., and Davidson, P.W.: Analysis of variance and F–tests for partial linear models with applications to environmental health data. *J. Amer. Statist. Assoc.*, 105, 991–1004 (2010)

24. Hušková, M., and Meintanis, S.G.: Omnibus tests for the error distribution in the linear regression model. *Statistics*, 41, 363–376 (2007)

25. Hušková, M., and Meintanis, S.G.: Tests for the error distribution in nonparametric possibly heteroscedastic regression models. *Test*, 19, 92–112 (2010)

26. Hušková, M., and Meintanis, S.G.: Tests for symmetric error distribution in linear and nonparametric regression models. *J. Statist. Plann. Infer.*, to appear (2011)

27. Jiménez–Gamero, M.D., Alba–Fernández, V., Muñoz–García, J., and Chalco–Cano, Y.: Goodness–of–fit tests based on the empirical characteristic function. *Comput. Statist. Dat. Anal.*, 53, 3957–3971 (2009)

28. Jurečková, J., Picek, J., and Sen, P.K.: Goodness-of-fit test with nuisance regression and scale. *Metrika*, 58, 235–258 (2003)

29. Klaassen, C.A. and Putter, H.: Efficient estimation of Banach parameters in semiparametric models. *Ann. Statist.*, 33, 307–346 (2005)

30. Lukacs, E.: *Developments in Characteristic Function Theory.* Griffin, London (1983)

31. Müller, U., Schick, A., and Wefelmeyer, W.: Estimating the error distribution in semiparametric regression. *Statistics and Decisions*, 25, 1–18 (2007)

32. Neumeyer, N.: Smooth residual bootstrap for empirical processes of nonparametric regression residuals. *Scand. J. Statist.*, 36, 204–228 (2009)

33. Neumeyer, N. and Dette, H.: Testing for symmetric error distribution in nonparametric regression models. *Statist. Sinica*, 17, 775–795 (2007)

34. Neumeyer, N., and van Keilegom, I.: Estimating the error distribution in nonparametric multiple regression with applications to model testing. *J. Multivar. Anal.*, 101, 1067–1078 (2010)

35. Neumeyer, N., Dette, H., and Nagel, E.R.: A note on testing symmetry of the error distribution in linear regression models. *J. Nonparam. Statist.*, 17, 697–715 (2005)

36. Neumeyer, N., Dette, H., and Nagel, E.R.: Bootstrap tests for the error distribution in linear and nonparametric regression models. *Austral. New Zeal. J. Statist.*, 48, 129–156 (2006)

37. Powell, B.: An introduction to smoothers. 4H Project Report for the degree of Master of Mathematics, Durham University, UK (2009)

38. Robinson, P.M.: Root-n Consistent Semiparametric Regression. *Econometrica*, 56, 931–954 (1988)

39. Ruppert, D., Wand, M.P., and Carroll, R.J.: *Semiparametric regression*. Cambridge Series in Statistical and Probabilist Mathematics. Cambridge University Press (2003)

40. Schick, A.: On asymptotically efficient estimation in semiparametric models. *Ann. Statist.*, 14, 1139–1151 (1986)

41. Schick, A.: Root–n–consistent and efficient estimation in semiparametric additive regression models. *Statist. Probab. Lett.*, 30, 45–51 (1996)

42. Sen, P.K., Jurečková, J., and Picek, J.: Goodness-of-fit test of Shapiro–Wilk type with nuisance regression and scale. *Austr. J. Statist.*, 32, 163–177 (2003)

43. Speckman, P.: Kernel smoothing in partial linear models. *J. Roy. Statist. Soc. B* , 50, 413–436 (1988)

44. Tenreiro, C.: On the choice of the smoothing parameter for the BHEP goodness–of–fit test. *Comput. Statist. Dat. Anal.*, 53, 1038–1053 (2009)

45. van der Vaart, A.W.: *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilist Mathematics. Cambridge University Press (1998)

46. Wahba, G.: Partial spline models for the semiparametric estimation of several variables. In: *Analyses for time series, Japan–US joint sem.*, pp. 319–329, Ames: Iowa States University Press (1984)